Abstract Title Page

Title:

Evaluation of Model Specification, Variable Selection, and Adjustment Methods in Relation to Propensity Scores and Prognostic Scores in Multilevel Data

Authors and Affiliations:

Bing Yu
Ontario Institute for Studies in Education of the University of Toronto

Guanglei Hong
University of Chicago

Background / Context:

How to effectively use propensity score methods (Rosenbaum & Rubin, 1983, 1984) in multilevel data has received increasing attention in the literature. The issues include how to select covariates for the propensity score model, how to make additional adjustment for covariates in the outcome model, and how to specify a multilevel propensity score model. The choice of an optimal analytic procedure may depend on how the multilevel data are generated. We consider three distinct multilevel settings representing different data generation processes. In a random intercept and slopes (RIS) setting, whether an individual will receive a treatment depends on individual characteristics, measured and unmeasured characteristics of the cluster to which this individual belongs, and certain interaction effects between the individual characteristics and the cluster characteristics. In a random intercept (RI) setting, unmeasured characteristics of clusters only affect the cluster average probability of treatment assignment. Individual treatment assignment may depend on individual characteristics, measured cluster characteristics, and their interactions. In a third setting, the treatment assignment occurs at the cluster level and hence does not depend on individual characteristics.

Because the data generation process is often unknown to the researchers, a user-specified multilevel propensity score model may or may not correspond to the data generation process. Previous research has examined the impacts on treatment effect estimation, comparing propensity score models that incorrectly represent the data generation structure with the correctly specified multilevel propensity score models. In the RIS setting, results from both simulation studies (Kelcey, 2009) and theoretical reasoning (Kim, 2007) have shown that multilevel propensity score models outperform single-level propensity score models in bias reduction. In the RI setting, when using within-cluster matching based on the estimated propensity scores, both random intercept propensity score models and single-level propensity score models would produce legitimate estimations of the propensity scores and would in turn lead to legitimate matches of the sample, as suggested by theoretical reasoning in Kim (2007). When using cross-cluster matching based on propensity scores estimated by multilevel models or by fixed-effect models, omitting the cluster-level information in the RI setting is not consequential for treatment effect estimation (Arpino, 2008). The impact of such omission in the RIS setting is yet to be examined. With regard to variable selection for propensity score models, past research has shown that including all the treatment predictors in a propensity score model increases variance without decreasing bias in comparison with including all the outcome predictors in a propensity score model, and thus the former approach is not recommended (Kelcey, 2009).

Alternative methods for additional adjustment through multilevel outcome models are yet to be studied in the propensity score literature. In the case of a binary outcome, analogous to a propensity score, a prognostic score is defined as the conditional probability of having a potential outcome value without the treatment, y_C, given the pretreatment covariates, $\mathbf{X}$, $\psi(\mathbf{X})=pr(y_C|\mathbf{X})$. Past research using single-level data has shown some advantageous properties of additional adjustment for prognostic scores (Hansen, 2006, 2008) or additional adjustment for covariates strongly predicting the outcome (Rubin & Thomas, 2000). This study tests whether such advantages are applicable in the multilevel settings. It is noteworthy that previous simulation studies have only used the RI structure for the outcome generation (Kelcey, 2009). In typical educational data, the outcome is as likely, if not more likely, to represent a RIS structure as does the treatment assignment. This study includes simulated outcomes with both the RIS structure and the RI structure.

Purpose / Objective / Research Question / Focus of Study:

This study uses simulation examples representing three types of treatment assignment mechanisms in data generation (the RIS setting, the RI setting, and a third setting with a cluster-level treatment and an individual-level outcome) in order to determine optimal procedures for reducing bias and improving precision in each of these three settings. Evaluation criteria include bias, variance, MSE, confidence interval coverage rate, and remaining sample size. Specifically, the study evaluates the performance of (a) three variable selection procedures for propensity score models (confounder-covariate PS models [CF], outcome-covariate PS models [OC], and treatment-covariate PS models [TC]), (b) three methods of adjustment through outcome models (adjusting for propensity scores [PS], adjusting for propensity scores in combination with prognostic scores [PS+Prog], and adjusting for propensity scores in combination with strong outcome predictors [PS+Cov]), and (c) four model specifications for the propensity score models and the prognostic score models (the RIS models with cross-level interactions [RIS Interaction], the RIS models without cross-level interactions [RIS], the RI models without cross-level interactions [RI], and the single-level models [Single]). For the last three types of model specification, the effect of omitting the cluster-level covariates is also under investigation (RIS-C, RI-C, Single-C). The RIS model specification will not be considered in the RI setting; the RIS and RI model specifications will not be considered in the third setting with cluster-level treatment assignment. Propensity scores and prognostic scores are adjusted through either non-parametric stratification or parametric covariance adjustment.

Significance / Novelty of study:

The study is among the first that considers simultaneously all three dimensions (i.e., variable selection for propensity score models, adjustment through outcome models, and propensity score model and prognostic score model specifications) in a variety of multilevel settings. For issues that have been approached in the past through theoretical reasoning only, the current study brings in additional dimensions for consideration and generates empirical evidence through simulations that complements the theoretical arguments and assists in reaching a comprehensive understanding. In particular, this study develops analytic procedures for prognostic score adjustment in multilevel data. Results from this study will provide practical guidance for applied researchers using propensity score methods in making causal inference.

Data Generation and Analysis:

In the RIS setting and RI setting, three cluster-level covariates, C_1, C_2, C_3, and three individual-level covariates, X_1, X_2, X_3, are independently generated from a standard normal distribution. C_1 and X_1 are true confounders which are related with both the treatment and the outcome, C_2 and X_2 are related with the outcome but not with the treatment, and C_3 and X_3 are related with the treatment but not with the outcome. There are *n* subjects nested in each of the *J* clusters.

In the RIS setting, the true propensity score is generated from a function specified as follows:

$$\text{logit}(P[Z=1|\mathbf{X}_{ij},\mathbf{C}_j]) = \alpha_{0j} + \alpha_{1j}X_{1ij} + \alpha_{2j}X_{2ij} + \alpha_{3j}X_{3ij}$$
$$\alpha_{0j} = \gamma_{00} + \gamma_{01}C_{1j} + \gamma_{02}C_{2j} + \gamma_{03}C_{3j} + \upsilon_{0j}$$
$$\alpha_{1j} = \gamma_{10} + \gamma_{11}C_{1j} + \gamma_{12}C_{2j} + \gamma_{13}C_{3j} + \upsilon_{1j}$$
$$\alpha_{2j} = \gamma_{20} + \gamma_{21}C_{1j} + \gamma_{22}C_{2j} + \gamma_{23}C_{3j} + \upsilon_{2j}$$

$$\alpha_{3j} = \gamma_{30} + \gamma_{31}C_{1j} + \gamma_{32}C_{2j} + \gamma_{33}C_{3j} + \upsilon_{3j}$$

$$\begin{pmatrix} \upsilon_{0j} \\ \upsilon_{1j} \\ \upsilon_{2j} \\ \upsilon_{3j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} \tau_{Z\,\text{intercept}} & 0 & 0 & 0 \\ 0 & \tau_{Z\,\text{slope}} & 0 & 0 \\ 0 & 0 & \tau_{Z\,\text{slope}} & 0 \\ 0 & 0 & 0 & \tau_{Z\,\text{slope}} \end{pmatrix}\right] = N\left[\mathbf{0} \quad \boldsymbol{\tau}_{\mathbf{Z}}\right]$$

The true potential outcomes are generated from a function specified as follows:

$$Y_{ij} = \delta Z + \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \beta_{3j}X_{3ij} + r_{ij}$$

$$\beta_{0j} = \pi_{00} + \pi_{01}C_{1j} + \pi_{02}C_{2j} + \pi_{03}C_{3j} + u_{0j}$$

$$\beta_{1j} = \pi_{10} + \pi_{11}C_{1j} + \pi_{12}C_{2j} + \pi_{13}C_{3j} + u_{1j}$$

$$\beta_{2j} = \pi_{20} + \pi_{21}C_{1j} + \pi_{22}C_{2j} + \pi_{23}C_{3j} + u_{2j}$$

$$\beta_{3j} = \pi_{30} + \pi_{31}C_{1j} + \pi_{32}C_{2j} + \pi_{33}C_{3j} + u_{3j}$$

$$r_{ij} \sim N\left[0 \quad \sigma^2\right],$$

$$\begin{pmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \\ u_{3j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} \tau_{Y\,\text{intercept}} & 0 & 0 & 0 \\ 0 & \tau_{Y\,\text{slope}} & 0 & 0 \\ 0 & 0 & \tau_{Y\,\text{slope}} & 0 \\ 0 & 0 & 0 & \tau_{Y\,\text{slope}} \end{pmatrix}\right] = N\left[\mathbf{0} \quad \boldsymbol{\tau}_{Y}\right]$$

In the RI setting, the true propensity score is generated from a function specified as follows:

$$\text{logit}(P[Z=1|\,\mathbf{X}_{ij}, \mathbf{C}_j]) = \alpha_0 + \alpha_1 X_{1ij} + \alpha_2 X_{2ij} + \alpha_3 X_{3ij} + \alpha_4 C_{1j} + \alpha_5 C_{2j} + \alpha_6 C_{3j} + \upsilon_{0j}$$

$$\upsilon_{0j} \sim N\left[0 \quad \tau_Z\right]$$

The true potential outcomes are generated from a function specified as follows:

$$Y_{ij} = \delta Z + \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \beta_3 X_{3ij} + \beta_4 C_{1j} + \beta_5 C_{2j} + \beta_6 C_{3j} + u_{0j} + r_{ij}$$

$$r_{ij} \sim N\left[0 \quad \sigma_Y^2\right], \qquad u_{0j} \sim N\left[0 \quad \tau_Y\right]$$

In the setting with a cluster-level treatment and an individual-level outcome, three cluster-level covariates, C_1, C_2, C_3, and one individual level covariate, X, are independently generated from a standard normal distribution. The true propensity score is generated from a function as follows:

$$\text{logit}(P[Z=1|\,\mathbf{X}_{ij}, \mathbf{C}_j]) = \alpha_{0j} + \alpha_{1j}C_{1j} + \alpha_{2j}C_{2j} + \alpha_{3j}C_{3j}$$

The potential outcomes are generated from a function as follows:

$$Y_{ij} = \delta_j Z + \beta_0 + \beta_1 C_{1j} + \beta_2 C_{2j} + \beta_3 C_{3j} + \beta_4 X_{ij} + u_{0j} + r_{ij}$$

$$r_{ij} \sim N\left[0 \quad \sigma_Y^2\right], \qquad u_{0j} \sim N\left[0 \quad \tau_Y\right]$$

Data are generated using parameter values that represent the typical values seen in educational research.

Findings / Results:

When examining bias reduction in the RIS setting, we have found that the largest variation exists between different methods for adjustment through the outcome model, with the

PS+Prog method outperforming the PS method and the PS+Cov method. The next important factor is specifications of propensity score models and prognostic score models. In the RIS setting, when using stratification, RIS models outperform less complex models (RI models or single-level models). When using covariance adjustment, RIS models (with or without interactions, with or without the cluster-level covariates) show advantages in minimizing MSE only when they are combined with prognostic score adjustment. Omitting the cluster-level covariates is not consequential for the RIS models and RI models. The single-level models do not perform well, with or without the cluster-level covariates. Variable selection for the propensity score models appears to be least consequential (please insert Tables 1 and 2 and Figure 1 here).

In the RI setting, in general, the covariance adjustment method has more stable performance than the stratification adjustment method. For covariance adjustment, all the alternative approaches under comparison lead to treatment effect estimates with minimal bias while the "PS + Prog" adjustment method has the smallest bias. For stratification adjustment, however, only the "PS + Prog" method has satisfactory performance. The "PS" method and the "PS + Cov" method both lead to largely biased estimates (please insert Tables 3 and 4 and Figure 2 here).

In the setting with a cluster-level treatment and an individual-level outcome, all alternative approaches are acceptable in terms of bias reduction. In general, the relative advantage of using the "PS + Prog" method lies primarily in precision improvement. However, when using stratification adjustment, additional adjustment for the prognostic score removes residual bias (please insert Table 5 and Figure 3 here).

Conclusions:

The findings from this study have implications for real data analysis. Before conducting data analysis, researchers should examine the multilevel data structure closely and collect additional information, if possible, on treatment assignment mechanisms. This is because the performance of alternative approaches differs across the three settings. Yet in all cases, propensity score adjustment in combination with prognostic score adjustment produces the optimal results. We reason that, in multilevel data, prognostic scores can effectively capture the multilevel structure in the outcome. Balancing on the prognostic scores therefore effectively increases the balance on the potential outcomes between the treatment groups.

Additionally, the multilevel propensity score models and multilevel prognostic score models do not always show advantages over their single-level counterparts. In the RIS setting, the advantages of using multilevel models only emerge when propensity score adjustment is combined with prognostic score adjustment or when stratification adjustment is used. In the RI setting, the performance of single-level models is very similar to that of multilevel models. Once the multilevel structure is accounted for through a random intercept outcome model, it matters very little whether the propensity scores and/or the prognostic scores included in the model are estimated through a multilevel model or a single level model. Omitting the cluster-level information does not appear to be consequential in either the RIS setting or the RI setting. Variable selection for propensity score models does not seem to matter much once the true confounders are included in the models.

Appendices

Appendix A. References

Arpino, B. & Mealli, F. (2008). The specification of the propensity score in multilevel observational studies. Unpublished MPRA paper.

Hansen, B. B. (2006). Bias reduction in observational studies via prognostic scores. Unpublished technical report.

Hansen, B. B. (2008). The prognostic analogue of the propensity score. *Biometrika*, *95*(2), 481-488.

Kelcey, B. M. (2009). Improving and assessing propensity score based causal inferences in multilevel and nonlinear settings. Unpublished doctoral dissertation. University of Michigan.

Kim, J. & Seltzer, M. (2007). Causal inference in multilevel settings in which selection process vary across schools. Unpublished technical report.

Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*, 41-55.

Rosenbaum, P. R. & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. Journal *of the American Statistical Association, 79*(387), 516-524.

Rubin, D. B. & Thomas, N. (2000). Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association, 95*, 573-585.

Appendix B. Tables and Figures

Table 1

Bias, Variance, MSE, 95% Confidence Interval Coverage Rate, and Sample Size of the Treatment Effect Estimates in the RIS Setting: Covariance Adjustment

Model type	Adjustment method	Variable selection method	Bias	Var.	MSE	CI CR	Sample size	Bias reduction %
RIS model with inter-actions	PS	TC	-3.920	0.244	0.685	0.85	5000	85.56
		OC	-6.458	0.225	0.825	0.74	5000	76.21
		CF	-4.398	0.210	0.668	0.81	5000	83.80
	PS + Prog	TC	0.181	0.146	0.286	0.96	5000	99.33
		OC	0.228	0.139	0.286	0.95	5000	99.16
		CF	0.172	0.133	0.282	0.95	5000	99.37
	PS + Cov	TC	-4.050	0.208	0.563	0.87	5000	85.08
		OC	-6.496	0.191	0.779	0.69	5000	76.07
		CF	-4.430	0.178	0.558	0.80	5000	83.68
RIS model	PS	TC	-3.239	0.243	0.636	0.88	5000	88.07
		OC	-5.462	0.226	0.704	0.83	5000	79.88
		CF	-3.202	0.211	0.579	0.87	5000	88.21
	PS + Prog	TC	0.464	0.144	0.289	0.95	5000	98.29
		OC	0.551	0.139	0.290	0.95	5000	97.97
		CF	0.489	0.133	0.287	0.94	5000	98.20
	PS + Cov	TC	-3.360	0.208	0.512	0.90	5000	87.62
		OC	-5.514	0.192	0.656	0.77	5000	79.69
		CF	-3.241	0.179	0.470	0.88	5000	88.06
RIS model without Cs	PS	TC	-3.607	0.245	0.660	0.86	5000	86.72
		OC	-5.778	0.226	0.738	0.81	5000	78.72
		CF	-3.635	0.211	0.608	0.86	5000	86.61
	PS + Prog	TC	0.263	0.146	0.291	0.95	5000	99.03
		OC	0.219	0.139	0.289	0.95	5000	99.19
		CF	0.243	0.134	0.287	0.95	5000	99.11
	PS + Cov	TC	-3.668	0.209	0.533	0.89	5000	86.49
		OC	-5.832	0.192	0.693	0.75	5000	78.52
		CF	-3.682	0.179	0.499	0.86	5000	86.44
RI model	PS	TC	3.846	0.216	0.651	0.83	5000	85.83
		OC	3.620	0.193	0.530	0.86	5000	86.66
		CF	3.623	0.193	0.598	0.82	5000	86.66
	PS + Prog	TC	4.326	0.165	0.566	0.77	5000	84.07
		OC	4.222	0.160	0.543	0.78	5000	84.45
		CF	4.213	0.160	0.542	0.78	5000	84.48
	PS + Cov	TC	3.688	0.184	0.531	0.83	5000	86.42
		OC	3.543	0.162	0.493	0.82	5000	86.95
		CF	3.548	0.162	0.493	0.82	5000	86.93

Table 1 (continued)

Model type	Adjustment method	Variable selection method	Bias	Var.	MSE	CI CR	Sample size	Bias reduction %
RI model without Cs	PS	TC	3.994	0.216	0.664	0.82	5000	85.29
		OC	3.738	0.193	0.539	0.85	5000	86.23
		CF	3.737	0.193	0.607	0.82	5000	86.24
	PS + Prog	TC	4.509	0.166	0.587	0.76	5000	83.39
		OC	4.332	0.160	0.555	0.77	5000	84.04
		CF	4.328	0.160	0.555	0.77	5000	84.06
	PS + Cov	TC	4.001	0.184	0.557	0.82	5000	85.26
		OC	3.777	0.162	0.510	0.81	5000	86.09
		CF	3.778	0.162	0.510	0.81	5000	86.09
Single level model	PS	TC	4.138	0.215	0.672	0.81	5000	84.76
		OC	3.972	0.191	0.554	0.84	5000	85.37
		CF	3.964	0.191	0.621	0.81	5000	85.40
	PS + Prog	TC	4.072	0.165	0.539	0.79	5000	85.00
		OC	3.902	0.158	0.511	0.80	5000	85.63
		CF	3.895	0.158	0.510	0.80	5000	85.66
	PS + Cov	TC	4.111	0.182	0.560	0.81	5000	84.86
		OC	3.909	0.158	0.512	0.80	5000	85.60
		CF	3.908	0.158	0.511	0.79	5000	85.61
Single level model without Cs	PS	TC	5.418	0.214	0.801	0.75	5000	80.04
		OC	5.148	0.191	0.669	0.76	5000	81.04
		CF	5.143	0.191	0.733	0.74	5000	81.06
	PS + Prog	TC	5.136	0.167	0.645	0.72	5000	81.08
		OC	4.970	0.161	0.615	0.73	5000	81.70
		CF	4.970	0.161	0.615	0.73	5000	81.70
	PS + Cov	TC	5.228	0.182	0.674	0.74	5000	80.74
		OC	4.97	0.161	0.615	0.73	5000	81.70
		CF	4.97	0.161	0.615	0.73	5000	81.70

Note: Estimation values of bias, variance, and MSE have been multiplied by 100 for ease of reading.

Table 2

Bias, Variance, MSE, 95% Confidence Interval Coverage Rate, and Sample Size of the Treatment Effect Estimates in the RIS Setting: Stratification Adjustment

Model type	Adjustment method	Variable selection method	Bias	Var.	MSE	CI CR	Sample size	Bias reduction %
RIS model with inter-actions	PS	TC	0.303	0.241	0.525	0.92	4759	98.88
		OC	-1.181	0.224	0.404	0.97	4774	95.65
		CF	1.480	0.211	0.503	0.90	4834	94.55
	PS + Prog	TC	0.259	0.167	0.335	0.95	4752	99.04
		OC	0.221	0.159	0.327	0.95	4768	99.19
		CF	0.413	0.153	0.327	0.93	4827	98.48
	PS + Cov	TC	0.162	0.206	0.396	0.96	4759	99.40
		OC	-1.232	0.191	0.365	0.95	4774	95.46
		CF	1.380	0.179	0.391	0.93	4834	94.92
RIS model	PS	TC	0.606	0.243	0.529	0.93	4766	97.77
		OC	-0.538	0.225	0.400	0.97	4785	98.02
		CF	2.168	0.213	0.531	0.88	4849	92.01
	PS + Prog	TC	0.328	0.167	0.341	0.95	4759	98.79
		OC	0.490	0.160	0.332	0.95	4778	98.20
		CF	0.637	0.153	0.335	0.93	4842	97.66
	PS + Cov	TC	0.466	0.207	0.398	0.96	4766	98.28
		OC	-0.605	0.192	0.361	0.96	4785	97.77
		CF	2.057	0.180	0.418	0.91	4849	92.42
RIS model without Cs	PS	TC	0.533	0.242	0.528	0.93	4719	98.04
		OC	-0.663	0.225	0.399	0.97	4745	97.56
		CF	2.000	0.212	0.521	0.89	4813	92.63
	PS + Prog	TC	0.638	0.168	0.341	0.95	4712	97.65
		OC	0.664	0.160	0.335	0.94	4738	97.55
		CF	0.895	0.153	0.337	0.93	4806	96.70
	PS + Cov	TC	0.473	0.207	0.399	0.95	4719	98.26
		OC	-0.643	0.191	0.359	0.96	4745	97.63
		CF	1.994	0.180	0.411	0.91	4813	92.65
RI model	PS	TC	6.452	0.220	0.922	0.70	4959	76.24
		OC	7.393	0.200	0.965	0.62	4956	72.77
		CF	7.407	0.200	1.034	0.61	4956	72.72
	PS + Prog	TC	1.689	0.174	0.415	0.90	4951	93.78
		OC	1.730	0.167	0.405	0.90	4948	93.63
		CF	1.694	0.167	0.404	0.90	4948	93.76
	PS + Cov	TC	6.267	0.188	0.792	0.68	4959	76.91
		OC	7.240	0.169	0.910	0.56	4956	73.33
		CF	7.219	0.169	0.906	0.57	4956	73.41

Table 2 (continued)

Model type	Adjustment method	Variable selection method	Bias	Var.	MSE	CI CR	Sample size	Bias reduction %
RI model without Cs	PS	TC	6.744	0.220	0.960	0.68	4953	75.16
		OC	7.700	0.200	1.012	0.59	4953	71.64
		CF	7.709	0.200	1.079	0.59	4953	71.60
	PS + Prog	TC	2.258	0.175	0.438	0.89	4945	91.68
		OC	2.267	0.167	0.430	0.89	4945	91.65
		CF	2.245	0.167	0.428	0.88	4945	91.73
	PS + Cov	TC	6.752	0.188	0.856	0.64	4953	75.13
		OC	7.811	0.169	0.996	0.52	4953	71.23
		CF	7.791	0.169	0.993	0.53	4953	71.31
Single level model	PS	TC	6.383	0.217	0.904	0.70	4978	76.49
		OC	6.268	0.195	0.804	0.69	4987	76.91
		CF	6.218	0.195	0.862	0.67	4987	77.10
	PS + Prog	TC	2.459	0.170	0.431	0.88	4969	90.94
		OC	3.077	0.162	0.458	0.85	4978	88.67
		CF	3.013	0.162	0.454	0.85	4978	88.90
	PS + Cov	TC	6.363	0.185	0.796	0.67	4978	76.56
		OC	6.280	0.162	0.768	0.62	4987	76.87
		CF	6.198	0.162	0.754	0.63	4987	77.17
Single level model without Cs	PS	TC	7.191	0.215	1.028	0.65	4986	73.52
		OC	6.981	0.194	0.899	0.63	4990	74.29
		CF	6.983	0.194	0.961	0.63	4990	74.28
	PS + Prog	TC	5.343	0.172	0.678	0.72	4978	80.32
		OC	5.478	0.166	0.679	0.69	4984	79.82
		CF	5.455	0.166	0.677	0.69	4984	79.91
	PS + Cov	TC	6.994	0.183	0.892	0.62	4986	74.24
		OC	6.832	0.164	0.842	0.59	4990	74.83
		CF	6.816	0.164	0.838	0.59	4990	74.90

Note: Estimation values of bias, variance, and MSE have been multiplied by 100 for ease of reading.

Table 3

Bias, Variance, MSE, 95% Confidence Interval Coverage Rate, and Sample Size of the Treatment Effect Estimates in the RI Setting: Covariance

Model type	Adjustment method	Variable selection method	Bias	Var.	MSE	CI CR	Sample size	Bias reduction %
RI model	PS	TC	-0.435	0.184	0.345	0.97	5000	98.17
		OC	-0.464	0.162	0.287	0.97	5000	98.05
		CF	-0.544	0.161	0.321	0.95	5000	97.71
	PS + Prog	TC	-0.013	0.131	0.254	0.96	5000	99.95
		OC	-0.002	0.127	0.249	0.96	5000	99.99
		CF	0.001	0.127	0.249	0.96	5000	100.00
	PS + Cov	TC	-0.472	0.151	0.277	0.97	5000	98.02
		OC	-0.494	0.129	0.255	0.95	5000	97.92
		CF	-0.490	0.129	0.255	0.95	5000	97.94
RI model without Cs	PS	TC	-0.320	0.184	0.344	0.97	5000	98.65
		OC	-0.373	0.162	0.286	0.97	5000	98.43
		CF	-0.457	0.161	0.321	0.95	5000	98.08
	PS + Prog	TC	0.111	0.131	0.257	0.96	5000	99.53
		OC	0.073	0.127	0.251	0.96	5000	99.69
		CF	0.076	0.127	0.251	0.96	5000	99.68
	PS + Cov	TC	-0.220	0.151	0.275	0.97	5000	99.08
		OC	-0.316	0.129	0.253	0.96	5000	98.67
		CF	-0.316	0.129	0.253	0.96	5000	98.67
Single level model	PS	TC	-0.138	0.183	0.342	0.97	5000	99.42
		OC	-0.114	0.160	0.285	0.97	5000	99.52
		CF	-0.183	0.160	0.316	0.96	5000	99.23
	PS + Prog	TC	-0.080	0.132	0.253	0.96	5000	99.67
		OC	-0.085	0.126	0.246	0.95	5000	99.64
		CF	-0.084	0.126	0.246	0.95	5000	99.65
	PS + Cov	TC	-0.069	0.149	0.273	0.97	5000	99.71
		OC	-0.086	0.126	0.247	0.95	5000	99.64
		CF	-0.085	0.126	0.247	0.95	5000	99.64
Single level model without Cs	PS	TC	0.956	0.183	0.350	0.96	5000	95.98
		OC	0.846	0.160	0.296	0.96	5000	96.44
		CF	0.777	0.160	0.322	0.95	5000	96.73
	PS + Prog	TC	0.732	0.133	0.262	0.95	5000	96.92
		OC	0.680	0.128	0.255	0.95	5000	97.14
		CF	0.680	0.128	0.255	0.95	5000	97.14
	PS + Cov	TC	0.839	0.150	0.282	0.96	5000	96.47
		OC	0.680	0.128	0.255	0.95	5000	97.14
		CF	0.680	0.128	0.255	0.95	5000	97.14

Note: Estimation values of bias, variance, and MSE have been multiplied by 10 for ease of reading.

Table 4

Bias, Variance, MSE, 95% Confidence Interval Coverage Rate, and Sample Size of the Treatment Effect Estimates in the RI Setting: Stratification Adjustment

Model type	Adjustment method	Variable selection method	Bias	Var.	MSE	CI CR	Sample size	Bias reduction %
RI model	PS	TC	3.625	0.187	0.483	0.89	4901	84.75
		OC	4.635	0.169	0.515	0.83	4919	80.50
		CF	4.578	0.168	0.540	0.81	4919	80.74
	PS + Prog	TC	0.110	0.145	0.284	0.97	4889	99.54
		OC	0.038	0.139	0.274	0.96	4909	99.84
		CF	0.032	0.139	0.275	0.96	4908	99.86
	PS + Cov	TC	3.568	0.153	0.412	0.87	4901	84.99
		OC	4.546	0.136	0.472	0.79	4919	80.88
		CF	4.533	0.136	0.470	0.79	4919	80.93
RI model without Cs	PS	TC	3.720	0.187	0.488	0.88	4898	84.35
		OC	4.690	0.168	0.520	0.83	4915	80.27
		CF	4.652	0.168	0.550	0.79	4915	80.43
	PS + Prog	TC	0.383	0.145	0.284	0.96	4886	98.39
		OC	0.330	0.139	0.277	0.95	4904	98.61
		CF	0.323	0.139	0.278	0.95	4904	98.64
	PS +Cov	TC	3.807	0.154	0.428	0.86	4898	83.99
		OC	4.807	0.136	0.497	0.77	4915	79.78
		CF	4.813	0.136	0.499	0.77	4915	79.76
Single level model	PS	TC	3.424	0.185	0.465	0.89	4951	85.60
		OC	3.609	0.165	0.425	0.88	4975	84.82
		CF	3.520	0.165	0.449	0.86	4976	85.19
	PS + Prog	TC	0.692	0.142	0.280	0.95	4938	97.09
		OC	1.164	0.136	0.279	0.94	4964	95.10
		CF	1.150	0.136	0.279	0.94	4964	95.16
	PS + Cov	TC	3.501	0.151	0.402	0.87	4951	85.27
		OC	3.682	0.131	0.393	0.83	4975	84.51
		CF	3.627	0.131	0.389	0.84	4976	84.74
Single level model without Cs	PS	TC	3.127	0.184	0.440	0.91	4978	86.85
		OC	3.023	0.163	0.387	0.91	4987	87.29
		CF	2.970	0.163	0.411	0.89	4987	87.51
	PS + Prog	TC	1.138	0.139	0.283	0.95	4969	95.21
		OC	1.367	0.134	0.282	0.93	4978	94.25
		CF	1.360	0.134	0.282	0.93	4978	94.28
	PS + Cov	TC	3.011	0.151	0.368	0.90	4978	87.33
		OC	2.886	0.131	0.342	0.88	4987	87.86
		CF	2.881	0.131	0.341	0.88	4987	87.88

Note: Estimation values of bias, variance, and MSE have been multiplied by 10 for ease of reading.

Table 5

Bias, Variance, MSE, 95% Confidence Interval Coverage Rate, and Sample Size of the Treatment Effect Estimates in the Setting with a Cluster-level Treatment and an Individual-level Outcome

Model type	Adjustment method	Variable selection method	Bias	Var.	MSE	CI CR	Sample size	Bias reduction %
Covariance Adjustment	PS	TC	0.179	0.665	1.173	0.97	12500	99.23
		OC	0.243	0.489	0.674	1.00	12500	98.96
		CF	0.168	0.486	0.980	0.95	12500	99.28
	PS + Prog	TC	0.146	0.177	0.354	0.94	12500	99.37
		OC	0.162	0.170	0.339	0.95	12500	99.31
		CF	0.159	0.170	0.339	0.95	12500	99.32
	PS + $X + C_2$	TC	0.153	0.336	0.515	0.99	12500	99.34
		OC	0.162	0.171	0.339	0.95	12500	99.30
		CF	0.162	0.171	0.339	0.95	12500	99.30
	PS+ X	TC	0.093	0.652	1.147	0.97	12500	99.60
		OC	0.163	0.477	0.646	1.00	12500	99.30
		CF	0.088	0.474	0.956	0.95	12500	99.62
Stratification Adjustment	PS	TC	1.602	0.685	1.240	0.97	12032	93.13
		OC	1.714	0.518	0.767	0.99	12166	92.65
		CF	1.625	0.516	1.063	0.94	12182	93.03
	PS + Prog	TC	0.374	0.206	0.405	0.95	12015	98.40
		OC	0.533	0.195	0.381	0.95	12152	97.71
		CF	0.499	0.195	0.390	0.94	12168	97.86
	PS + $X + C_2$	TC	1.583	0.352	0.576	0.98	12032	93.21
		OC	1.680	0.197	0.422	0.93	12166	92.80
		CF	1.610	0.196	0.418	0.94	12182	93.10
	PS+ X	TC	1.602	0.685	1.240	0.97	12032	93.13
		OC	1.714	0.518	0.767	0.99	12166	92.65
		CF	1.625	0.516	1.063	0.94	12182	93.03

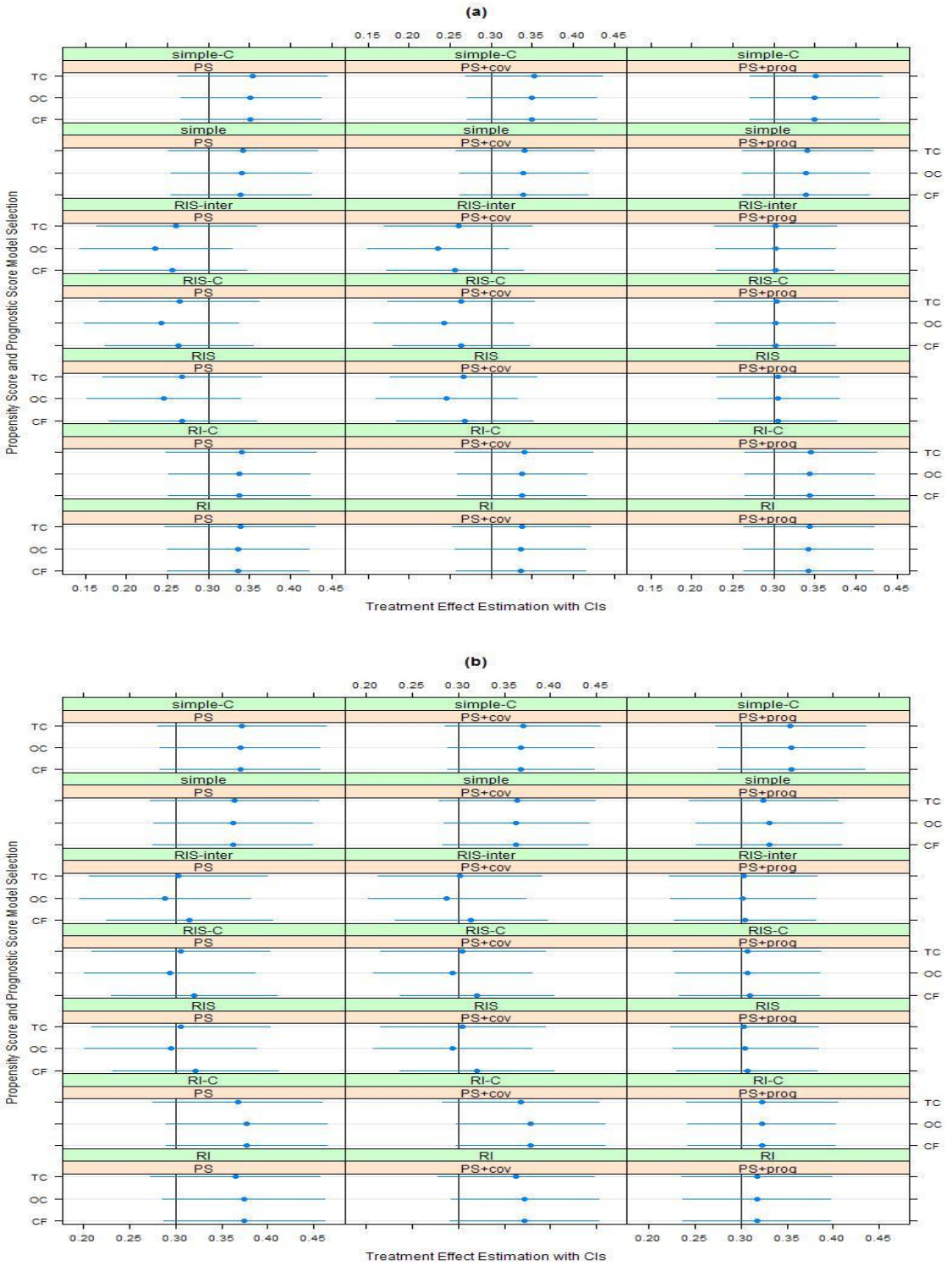

Figure 1. **Treatment Effect Estimations with 95% Confidence Intervals in the RIS Setting: Covariance Adjustment (a) and Stratification Adjustment (b).**

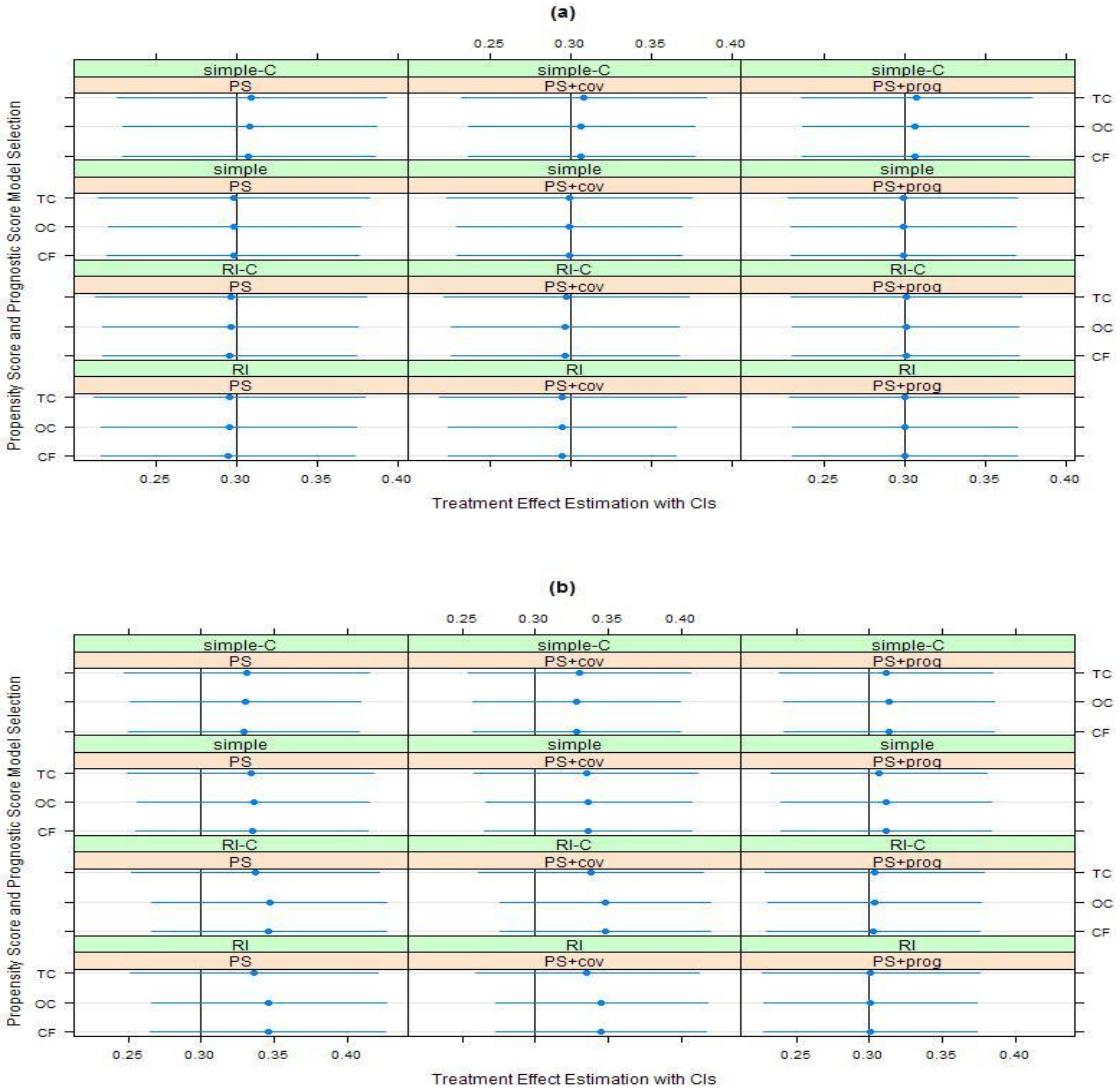

Figure 2. **Treatment Effect Estimations with 95% Confidence Intervals in the RI Setting: Covariance Adjustment (a) and Stratification Adjustment (b).**

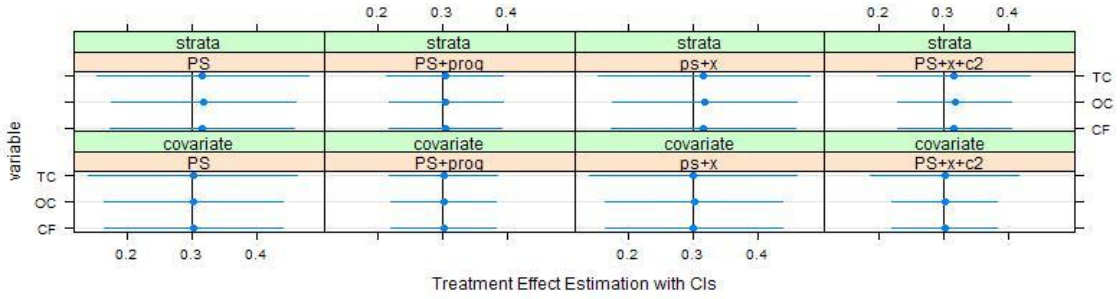

Figure 3. **Treatment Effect Estimations with 95% Confidence Intervals in the Setting with a Cluster-level Treatment and an Individual-level Outcome.**